

## VU Research Portal

### **De praktijk van systematische reviews: IX Plaatsbepaling in behandelbeslissingen en richtlijnen**

Assendelft, W.J.J.; van Tulder, M.W.; van Everdingen, J.J.; Offringa, M.; Bouter, L.M.

#### ***published in***

Nederlands Tijdschrift voor Geneeskunde  
2002

#### ***document version***

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

#### ***citation for published version (APA)***

Assendelft, W. J. J., van Tulder, M. W., van Everdingen, J. J., Offringa, M., & Bouter, L. M. (2002). De praktijk van systematische reviews: IX Plaatsbepaling in behandelbeslissingen en richtlijnen. *Nederlands Tijdschrift voor Geneeskunde*, 146, 1527-1531.

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

#### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

- 22 Dye C, Williams BG, Espinal MA, Raviglione MC. Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. *Science* 2002;295:2042-6.
- 23 Blower SM, Gerberding JL. Understanding, predicting and controlling the emergence of drug-resistant tuberculosis: a theoretical framework. *J Mol Med* 1998;76:624-36.
- 24 Dye C, Williams BG. Criteria for the control of drug-resistant tuberculosis. *Proc Natl Acad Sci USA* 2000;97:8180-5.
- 25 Tahaoglu K, Torun T, Sevim T, Atac G, Kir A, Karasulu L, et al. The treatment of multidrug-resistant tuberculosis in Turkey. *N Engl J Med* 2001;345:170-4.
- 26 Iseman MD, Madsen L, Goble M, Pomerantz M. Surgical intervention in the treatment of pulmonary disease caused by drug-resistant *Mycobacterium tuberculosis*. *Am Rev Respir Dis* 1990;141:623-5.
- 27 Geerligs WA, Altena R van, Lange WCM de, Soolingen D van, Werf TS van der. Multidrug-resistant tuberculosis: long-term treatment outcome in the Netherlands. *Int J Tuberc Lung Dis* 2000;4:758-64.
- 28 Espinal MA, Kim SJ, Suarez PG, Kam KM, Khomenko AG, Migliori GB, et al. Standard short-course chemotherapy for drug-resistant tuberculosis: treatment outcomes in 6 countries. *JAMA* 2000;283:2537-45.
- 29 Farmer P, Bayona J, Becerra M, Furin J, Henry C, Hiatt H, et al. The dilemma of MDR-TB in the global era [letter]. *Int J Tuberc Lung Dis* 1998;2:869-76.
- 30 Espinal MA, Dye C, Raviglione M, Kochi A. Rational 'DOTS plus' for the control of MDR-TB. *Int J Tuberc Lung Dis* 1999;3:561-3.
- 31 Farmer P, Leandre F, Mukherjee JS, Gupta R, Tarter L, Kim JY. Community-based treatment of advanced HIV disease: introducing DOT-HAART (directly observed therapy with highly active antiretroviral therapy). *Bull World Health Organ* 2001;79:1145-51.
- 32 Mbewu A, Feachem R, Pablos-Mendez A, Gilks C, AbouZhar C, Turmen T. Antiretroviral therapy is only part of it [letter]. *Bull World Health Organ* 2001;79:1152-5.

Aanvaard op 17 juni 2002

## Capita selecta

# De praktijk van systematische reviews. IX. Plaatsbepaling in behandelbeslissingen in richtlijnen

W.J.J.ASSENDELFT, M.W.VAN TULDER, J.J.E.VAN EVERDINGEN, M.OFFRINGA EN L.M.BOUTER

In een systematische review (SR) wordt voor het beantwoorden van een klinische vraag het klinische bewijs uit verschillende studies systematisch en zo volledig mogelijk verzameld; vervolgens wordt dit beoordeeld op de mogelijkheid van vertekening (bias) en op reproduceerbare wijze samengevoegd.<sup>1</sup> De resultaten van een systematische review worden gebruikt in de individuele patiëntenzorg of bij het opstellen van een behandelrichtlijn (figuur 1).

Bij het nemen van een behandelbeslissing moet rekening gehouden worden met de verschillende factoren die de kracht van het bewijs bepalen: het studietype, de consistentie van de resultaten (indien meerdere studies beschikbaar zijn), de kwaliteit van de afzonderlijke studies, de grootte van het effect, de precisie waarmee een effect is vastgesteld en de klinische relevantie van de gebruikte uitkomstmaten.<sup>2-4</sup>

## SAMENVATTING

– In een systematische review (SR) wordt omtrent een klinische vraagstelling het klinische bewijs uit verschillende studies systematisch en zo volledig mogelijk verzameld, op de mogelijkheid voor vertekening (bias) beoordeeld en op reproduceerbare wijze samengevoegd.

– De resultaten van een SR kunnen worden gebruikt in de individuele patiëntenzorg of bij het opstellen van een behandelrichtlijn.

– Een vast systeem volgens hetwelk het beschikbare bewijs naar validiteit van het gebezigde studietype wordt geordend (bewijsniveau of 'levels of evidence') wordt overal in de wereld toegepast door (nationale) organisaties voor richtlijnontwikkeling, ook in Nederland.

– Bij het nemen van een behandelbeslissing moet rekening gehouden worden met de verschillende factoren die de kracht van het bewijs bepalen: het studietype, de consistentie van de resultaten (indien meerdere studies beschikbaar zijn), de kwaliteit van de afzonderlijke studies, de grootte van het effect, de precisie waarmee een effect is vastgesteld, en de klinische relevantie van de gebruikte uitkomstmaten. Binnen een SR kan dit problemen opleveren omdat er soms onvoldoende studies met dezelfde uitkomsten zijn.

– Ook blijkt in de praktijk de inschatting van de mate van bewijskracht van een bepaalde SR ten opzichte van een andere review of primaire studies problemen op te leveren. Er kunnen over hetzelfde onderwerp tegenstrijdige reviews zijn, er kunnen verschillen bestaan tussen reviews en megatrials of er kan maar 1 gerandomiseerde trial beschikbaar zijn. Deze problemen kunnen vaak opgelost worden met een systematische analyse. Dit vereist in de regel zowel methodologische als vakinhoudelijke kennis.

Academisch Medisch Centrum/Universiteit van Amsterdam, Amsterdam.

Dutch Cochrane Centre en Divisie Public Health, Afd. Huisartsgeneeskunde: dr.W.J.J.Assendelft, huisarts-epidemioloog (thans: Nederlands Huisartsen Genootschap, Postbus 3231, 3502 GE Utrecht).

Emma Kinderziekenhuis, afd. Kindergeneeskunde: prof.dr.M.Offringa, kinderarts-epidemioloog (tevens: Dutch Cochrane Centre).

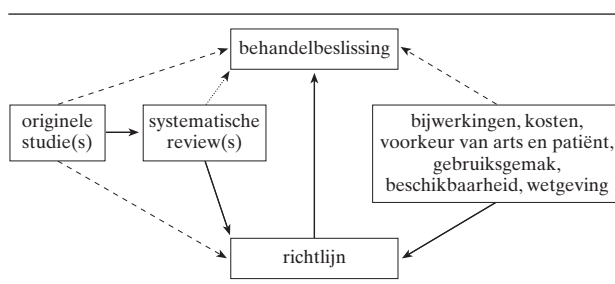
Vrije Universiteit, Instituut voor Extramuraal Geneeskundig Onderzoek, Amsterdam.

Dr.M.W.van Tulder en prof.dr.L.M.Bouter, epidemiologen.

Kwaliteitsinstituut voor de Gezondheidszorg CBO, Utrecht.

Dr.J.J.E.van Everdingen, adjunct-directeur medisch-specialistische kwaliteit.

Correspondentieadres: dr.W.J.J.Assendelft (p.assendelft@nhg-nl.org).



FIGUUR 1. De plaats van de systematische review (SR) in een behandelbeslissing; (—): de meest efficiënte en transparante manier van het gebruik van een SR; (---): een minder efficiënte mogelijkheid; (.....): een minder geschikte mogelijkheid.

In dit artikel zullen wij bespreken welke rol deze factoren spelen in de behandelbeslissing. Problemen bij het toepassen ervan binnen een SR worden toegelicht. Daarbij gaan wij ook in op SR's waarin vanwege ontbrekende gegevens of een te grote heterogeniteit niet statistisch kan worden gepoold. Tenslotte gaan wij in op de praktische problemen die optreden bij gebruik van bewijs uit SR's in een behandelbeslissing. De andere aspecten van een behandelbeslissing (bijwerkingen, kosten, voorkeuren van arts en patiënt, beschikbaarheid en juridische randvoorwaarden) staan elders beschreven.<sup>3-5</sup> Om reden van overzichtelijkheid beperken wij ons tot SR's van gerandomiseerde effectstudies (gerandomiseerde klinische trials; RCT's). Veel aspecten zijn echter ook van toepassing op SR's van observationele studies.<sup>6,7</sup>

Een SR dient uit te gaan van een expliciete klinische vraag. De noodzakelijke ingrediënten van een dergelijke vraag zijn: de patiëntencategorie, de interventie, de controle-interventie, de uitkomst en eventueel ook de gewenste follow-upduur. Het heeft pas zin om naar de validiteit, de uitkomsten en de toepasbaarheid van die uitkomsten van een SR te kijken als deze voldoende aansluit op de klinische vraag.

#### DIMENSIES VAN BEWIJS

**Studiotype.** Het is voor de richtlijnontwikkelaar en de behandelend arts nuttig om snel inzicht te hebben in de mate van bewijskracht van het aan de behandelbeslissing ten grondslag liggende onderzoek. Een vast systeem volgens welke het beschikbare bewijs naar validiteit van het gebezigde studiotype wordt geordend (niveaus van bewijs of 'levels of evidence') werd het eerst toegepast in Canada in 1979.<sup>8</sup> Momenteel wordt een dergelijk systeem overall in de wereld toegepast door (nationale) organisaties voor richtlijnontwikkeling, ook in Nederland. Het in de tabel gepresenteerde systeem van de Scottish Intercollegiate Guidelines Network (SIGN) ([www.sign.ac.uk](http://www.sign.ac.uk)) verschilt op details van het huidige systeem van het Kwaliteitsinstituut voor de Gezondheidszorg CBO.<sup>9</sup>

Dergelijke bewijsniveausystemen bestaan voor effectiviteitsonderzoek, maar ook voor bijvoorbeeld diagnostiek. Het maken van richtlijnen volgens een bewijsni-

veausysteem zoals gepresenteerd in de tabel is het eenvoudigst als er voldoende, informatief en hoogwaardig onderzoek beschikbaar is dat onderling niet in tegenspraak is. De praktijk is echter vaak weerbarstig.

**Kwaliteit van de studies.** Indien een SR de basis van een aanbeveling vormt, hangt de zeggingskracht allereerst samen met de kwaliteit van de SR zelf. Een eerder artikel in deze reeks geeft hiervoor duidelijke richtlijnen.<sup>10</sup> Belangrijke onderdelen van de kwaliteit van een SR zijn de volledigheid van de zoekactie, de beoordeling van de validiteit van de primaire studies en een inzichtelijke presentatie van de uitkomsten.<sup>10</sup> Hiernaast dient ook een indruk van de validiteit van de in de SR opgenomen studies te worden gevormd.<sup>11</sup> Een SR gebaseerd op valide primaire studies heeft vanzelfsprekend meer zeggingskracht dan een SR gebaseerd op primaire studies van dubieuze validiteit. Het beoordelen van kwaliteit is echter niet eenvoudig: er zijn veel kwaliteitsbeoordelingslijsten beschikbaar, die ook nog eens sterke onderlinge verschillen in scores kunnen geven.<sup>12</sup> Van een aantal validiteitsitems staat echter vast dat het niet correct uitvoeren kan leiden tot vertekening van de uitkomst van de studie: blinding van de randomisatie en blinding van patiënten en behandelaars.<sup>11</sup> Deze items kunnen dus in een onderlinge rangschikking van studies worden gebruikt.

Indien een goede SR ontbreekt, moet de behandelbeslissing worden gebaseerd op primaire studies die eveneens op kwaliteit moeten worden beoordeeld.

**Grootte, precisie en klinische relevantie van het effect.** Bij een SR met statistische pooling betreft dit de gepoolde effectmaat. De resultaten van effectstudies kunnen op twee manieren worden weergegeven. Dichotome variabelen (genezen-ziek blijven; overleven-overlijden) worden meestal uitgedrukt in een oddsratio (OR), een relatief risico (RR) of een risicoverschil (RV). Continue

Voorbeeld van een systeem om gegevens te ordenen die de onderbouwing vormen van een diagnostische of therapeutische beslissing of aanbeveling (Scottish Intercollegiate Guidelines Network, Schotland; [www.sign.ac.uk](http://www.sign.ac.uk))\*

- 1++ hoge kwaliteit SR's van RCT's, of RCT's met een zeer lage kans op bias
- 1+ goed uitgevoerde SR's van RCT's, of RCT's met een lage kans op bias
- 1- SR's van RCT's, of RCT's met een hoge kans op bias
- 2++ hoge kwaliteit SR's van patiënt-controle- of cohortstudies; hoge kwaliteit patiënt-controle- of cohortstudies met een zeer lage kans op vertekening ('confounding'), bias of toevalsfouten en een grote waarschijnlijkheid dat de relatie causaal is
- 2+ goed uitgevoerde patiënt-controle- of cohortstudies met een lage kans op confounding, bias of toevalsfouten en een redelijke waarschijnlijkheid dat de relatie causaal is
- 2- patiënt-controle- of cohortstudies met een hoge kans op confounding, bias of toevalsfouten en een aanzienlijke kans dat de relatie niet causaal is
- 3 niet vergelijkende studies, bijvoorbeeld groepen patiënten of gevalsbeschrijvingen
- 4 deskundigenoordeel; geen ondersteunende studies aanwezig

SR = systematische review; RCT = gerandomiseerde klinische trial.

\*Dit systeem verschilt op details van het huidige Nederlandse systeem van het Kwaliteitsinstituut voor de Gezondheidszorg CBO.<sup>9</sup>

variabelen worden uitgedrukt in een verschilscore, meestal het verschil tussen de gemiddelden van de groepen. Bij de uitkomsten wordt een 95%-betrouwbaarheidsinterval (95%-BI) gerapporteerd. Een (gepooled) resultaat is statistisch significant indien het 95%-BI bij een OR of RR de 'neutrale waarde' 1 uitsluit (dan is er geen verschil tussen de interventie- en de controlegroep) of indien, bij een RV of een continue uitkomst (verschilscore), het 95%-BI de 'neutrale waarde' 0 (geen verschil) uitsluit.<sup>10</sup>

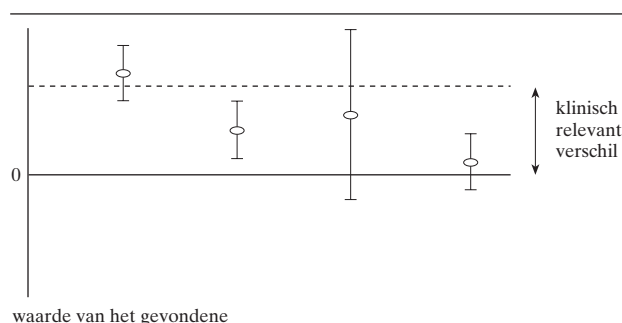
Een statistisch significant verschil betekent echter niet dat dit effect in de praktijk ook de moeite waard is. Het effect kan klein zijn en de inschatting van de klinische relevantie van de omvang van het verschil wordt dan ook mede bepaald door de afweging tegen de kosten en de bijwerkingen.<sup>13 14</sup> In figuur 2 wordt de relatie tussen grootte, precisie en klinische relevantie van het effect weergegeven.<sup>15</sup> Uit deze figuur wordt duidelijk dat een grote studie bijvoorbeeld statistisch significante resultaten kan hebben, maar dat de uitkomst in de praktijk niet relevant hoeft te zijn. Andersom kan het ook zo zijn dat het gevonden verschil weliswaar klinisch relevant is, maar dat de studie niet groot genoeg is (niet genoeg onderscheidingsvermogen ('power') heeft) om dit verschil ook statistisch significant te laten zijn.

**Relevantie van de uitkomstmaat.** Vaak zijn er verschillende uitkomstmaten beschikbaar. Idealiter dient een richtlijn gebaseerd te zijn op evident relevante klinische uitkomsten (zoals overleving, fractuurincidentie en wondheling) of op subjectieve patiëntgeoriënteerde uitkomstmaten (zoals dagelijks functioneren of kwaliteit van leven).<sup>16</sup> Te vaak gebeurt het dat in een onderzoek slechts gebruik wordt gemaakt van surrogaatuitkomstmaten. Het is een probleem wanneer de in de SR opgenomen RCT's in meerderheid slechts surrogaatuitkomsten (serumcholesterolconcentratie, botmineraaldichtheid) en niet de 'harde' klinische uitkomst (myocardinfarct, femurfractuur) rapporteren. Vaak gebeurt dit omdat de studieomvang te gering of de studieduur te kort is om klinische uitkomstmaten te kunnen rapporteren. Ook kan het zijn dat men zich in een studie tot een surrogaatuitkomst heeft beperkt omdat in andere (interventie)studies de relatie tussen de surrogaatuitkomst en de relevantere klinische uitkomst reeds is aangetoond. De waarde van een surrogaatuitkomst neemt toe indien aan meer van de volgende criteria wordt voldaan: er is een aangetoond sterk en consistent verband met een klinische of patiëntgeoriënteerde uitkomst en de rol van de surrogaatuitkomst is biologisch en pathofysiologisch plausibel.<sup>2 17</sup>

#### INTERPRETATIE VAN BEWIJSKRACHT BIJ

##### ONVOLDOENDE STUDIES MET GELIJKE UITKOMSTEN

Het is een voorwaarde voor statistisch poolen dat de studies klinisch voldoende gelijk ('homogeen') zijn.<sup>18</sup> Het komt echter vaak voor dat de studies weliswaar klinisch homogeen zijn, maar dat het onmogelijk is om te poolen doordat er statistische heterogeniteit bestaat, de uitkomstmaten onderling te verschillend zijn, of doordat de gerapporteerde gegevens onvoldoende gedetailleerd



waarde van het gevondene

relevantant significant	ja ja	nee ja	? nee	nee nee
bruikbaar	ja	nee	onduidelijk	nee

FIGUUR 2. Visualisatie van het onderscheid tussen klinische relevantie en statistische significantie.<sup>15</sup> De verticale lijnen geven de 95%-betrouwbaarheidsintervallen rond het verschil tussen behandeling en controle weer.

zijn om in een pooling te worden gebruikt (bijvoorbeeld omdat een standaarddeviatie ontbreekt). Toch is er dan de behoefte om op objectieve en reproduceerbare wijze tot een samenvattende uitspraak te komen over de bewijskracht van de SR. Het simpelweg optellen van positieve studies enerzijds en negatieve anderzijds is dan onvoldoende, omdat dan slechts recht wordt gedaan aan één van de dimensies van bewijskracht, namelijk de statistische significantie. Het is duidelijk beter wanneer naast statistische significantie in een afweging ook de klinische relevantie en consistentie van de studies wordt betrokken. Het toepassen van deze dimensies in een tevoren geconstrueerd algoritme maakt het proces van afwegen transparanter en reproduceerbaarder. De getalsmatige afwegingen in dergelijke algoritmen zijn echter arbitrair en kunnen slechts ter ondersteuning van een beslissing dienen, die verder vooral ook op klinische interpretatie van de resultaten is gebaseerd. Ook de constatering dat de studies te heterogeen zijn om tot een samenvattend oordeel te kunnen komen, behoort tot de mogelijke conclusies.

De invloed van de kwaliteit van de individuele studies kan met een sensitiviteitsanalyse worden geanalyseerd. De synthese wordt gemaakt met allereerst alle studies en vervolgens alleen de studies van voldoende hoge kwaliteit. Indien de laatste analyse een ander resultaat geeft, dan zal dit in de SR moeten worden aangegeven.<sup>11</sup>

#### SYSTEMATISCHE REVIEW EN BEHANDELBESSLINGEN

In de praktijk blijkt de inschatting van de mate van bewijskracht van een bepaalde SR ten opzichte van een andere review of primaire studies vaak problemen op te leveren. Deze problemen kunnen echter vaak weer op systematische wijze geanalyseerd en opgelost worden. Dit vereist in de regel zowel methodologische als vakinhoudelijke kennis.

**Tegenstrijdige reviews.** Het is mogelijk dat er over dezelfde klinische vraagstelling SR's zijn met resultaten die

met elkaar in tegenspraak lijken te zijn.<sup>19-21</sup> Een dergelijke tegenstrijdigheid kan in de regel met gebruikmaking van een systematische analyse van de verschillen in vraagstelling en uitvoering van de SR's inzichtelijk worden gemaakt.<sup>10-22</sup>

*Tegenstrijdigheid tussen review en megatrial.* Soms is er over een onderwerp zowel een SR van kleine studies als een grote RCT (megatrial) beschikbaar, die met elkaar in tegenspraak zijn. Dit is een reëel methodologisch probleem en het zal vaak moeilijk zijn een keuze te maken.<sup>23-24</sup> In sommige systemen voor het bepalen van de bewijskracht (bewijsniveau) heeft één enkele megatrial een gelijkwaardige plaats aan de SR (zie de tabel), terwijl in andere systemen de SR een hogere plaats krijgt (<http://cebm.jr2.ox.ac.uk>).<sup>9</sup> De keuze zal deels moeten worden bepaald door de klinische vraagstelling van de richtlijnontwikkelaars: zijn er verschillen in geïncludeerde patiënten, toegepaste interventies en uitkomstmaten, en welke van beide (SR of megatrial) sluit het best aan bij de klinische vraagstelling? Daarnaast zal bij de keuze vanzelfsprekend ook de kwaliteit van beide studietypen een rol moeten spelen. Tenslotte is bij een SR van kleine RCT's met name publicatiebias een probleem: kleinere positieve RCT's hebben een grotere kans om gepubliceerd te worden dan kleinere negatieve studies, hetgeen tot een overschatting van het effect van een interventie in de SR zal leiden. Er bestaan statistische en grafische technieken om de aanwezigheid van publicatiebias te onderzoeken.<sup>25</sup>

*Plaatsbepaling van slechts één megatrial.* Soms is er slechts 1 megatrial met klinisch relevante uitkomsten beschikbaar, die door velen in de praktijk reeds als doorslaggevend bewijs wordt beschouwd. Zoals eerder opgemerkt, wordt het wetenschappelijk bewijs van een enkele megatrial in richtlijnen doorgaans op een minder hoog niveau gezet dan een SR van meerdere RCT's (<http://cebm.jr2.ox.ac.uk>).<sup>9</sup> Allereerst omdat een positief resultaat op statistische gronden een fout-positieve uitkomst kan zijn (bij een significantieniveau van 0,05 is 1 op de 20 studies fout-positief). Daarnaast is een effect dat in meerdere RCT's onder verschillende omstandigheden (landen, soorten gezondheidszorgsystemen, echelons) is vastgesteld, beter generaliseerbaar dan resultaten uit slechts 1 megatrial.

*Ontbreken van RCT's.* Een RCT, en derhalve ook een SR van RCT's, over de effectiviteit van een bepaalde interventie is soms niet wenselijk, mogelijk of nog niet beschikbaar.<sup>26</sup> In dat geval zal men genoegen moeten nemen met (een SR van) observationele studies. Hierbij zal men zich moeten realiseren dat observationele studies gevoeliger zijn voor systematische vertekening (bias).<sup>6-27</sup> In veel gevallen geven observationele studies echter dezelfde uitkomsten als RCT's; het is dus onjuist om in afwezigheid van RCT's de resultaten van observationele studies volledig te negeren.<sup>28-30</sup>

#### BESCHOUWING

Steeds vaker wordt in een behandelrichtlijn de bewijskracht van het onderliggende onderzoek in de presentatie meegenomen. Vanwege de grondigheid, de volledigheid

en de transparantie is een SR een goede basis voor een richtlijn, en krijgt derhalve het hoogste bewijsniveau. Indien geen SR's voorhanden zijn, dan kan gebruikgemaakt worden van resultaten van (gerandomiseerde) effectstudies, en wanneer die ontbreken, van observationele studies. De bewijskracht neemt echter met ieder niveau af, en de sterkte van de aanbeveling zal met een lager niveau van bewijskracht zelf ook minder zijn. De referenties van de individuele studies die aan de richtlijn hebben bijgedragen, moeten voor de gebruiker toegankelijk zijn ([www.agreecollaboration.org](http://www.agreecollaboration.org)). Om reden van overzichtelijkheid gebeurt dit steeds vaker in een apart begeleidend document.

Er zijn echter meer factoren dan alleen het wetenschappelijk bewijs die een behandeladvies bepalen. Het bewijsniveau gebaseerd op het studietype is belangrijk, maar het vormt zeker niet de enige component van een behandeladvies. Daarnaast vergt het interpreteren van onvolkomen of tegenstrijdig bewijs een grondig inzicht in de klinische epidemiologie en in het betreffende onderwerp. Het zonder methodologische en vakinhoudelijke achtergrond direct vertalen van de resultaten van een SR naar een behandeladvies leidt dan ook gemakkelijk tot verkeerde conclusies. Het wetenschappelijk bewijs dient echter wel als basis voor aanbevelingen en een expliciete, geordende weergave is derhalve een noodzakelijke voorwaarde voor een behandelrichtlijn. Richtlijnontwikkelaars dienen bij belangrijke of controversiële onderwerpen te overwegen of het gewenst is om een SR te (laten) ontwikkelen van het bewijsmateriaal uit beschikbare afzonderlijke studies.<sup>31-32</sup> Tenslotte dient men zich te realiseren dat klinisch bewijs of een richtlijn onvoldoende is om de dagelijkse praktijk te veranderen.<sup>33-34</sup> Aan implementatie zal dus aparte aandacht moeten worden gegeven.<sup>34</sup>

Het systeem in Nederland, waarbij beroepsgroepen – ondersteund door de overheid – zelf richtlijnen maken, sluit hier goed bij aan. Omdat de inschatting en de afweging van dimensies van bewijskracht in de praktijk vaak problemen opleveren, is de beschikbaarheid van methodologische expertise in richtlijncommissies van groot belang. De ontwikkeling van ideeën over dimensies van bewijskracht is nog lang niet ten einde. Toch is het voor de eindgebruiker prettig indien hij bij richtlijnen telkens hetzelfde systeem van kwaliteitsbeoordeling van studies en dezelfde bewijsniveaus tegenkomt. Dit zal de richtlijnen inzichtelijker maken en de acceptatie en het gebruiksgemak ervan verhogen.

#### ABSTRACT

*Systematic reviews in practice. IX. Determining their place in treatment recommendations in practice guidelines*

– In a systematic review (SR), the available evidence for a clinical problem is systematically and comprehensively collected from different studies, the likelihood of bias is assessed and the results are summarised in a reproducible manner.

– The results of an SR can be used in either individual patient care or the formulation of a practice guideline.

– Throughout the world, (national) organisations for guideline development use systems for classifying the validity of evidence



according to the study design ('levels of evidence'), and this is also the case in the Netherlands.

– In formulating a treatment recommendation, consideration must be given to the various factors which determine the strength of the evidence: study design, consistency of the results (if more studies are available), quality of the individual studies, magnitude and precision of the reported effect and the clinical relevance of the outcome measures. If insufficient studies with the same outcome measure are available, problems can arise in an SR.

– In practice, problems can also arise with respect to estimating the cogency of an SR compared to another review or primary studies. For the same subject, there might be discordant reviews, differences between reviews and megatrials might exist or just one randomised clinical trial might be available. These problems can often be solved with the aid of a systematic analysis. This requires both methodological and clinical expertise.

#### LITERATUUR

- 1 Offringa M, Craen AJM de. De praktijk van systematische reviews. I. Inleiding. Ned Tijdschr Geneesk 1999;143:653-6.
- 2 O'Connell D, Glasziou P, Hill S. How to use the evidence: assessment and application of scientific evidence. Canberra: National Health and Medical Research Council; 2000.
- 3 Guyatt GH, Sinclair J, Cook DJ, Glasziou P. Users' guides to the medical literature. XVI. How to use a treatment recommendation. Evidence-Based Medicine Working Group and the Cochrane Applicability Methods Working Group. JAMA 1999;281:1836-43.
- 4 Sheldon TA, Guyatt GH, Haines A. Getting research findings into practice. When to act on the evidence. BMJ 1998;317:139-42.
- 5 Haynes B, Haines A. Barriers and bridges to evidence based clinical practice. BMJ 1998;317:273-6.
- 6 Windt DAWM van der, Zeegers MPA, Kemper HCG, Assendelft WJJ, Scholten RJPM. De praktijk van systematische reviews. VI. Zoeken, selecteren en methodologisch beoordelen van etiologisch onderzoek. Ned Tijdschr Geneesk 2000;144:1210-4.
- 7 Zeegers MP, Heisterkamp SH, Kostense PJ, Windt DAWM van der, Scholten RJPM. De praktijk van systematische reviews. VII. Het combineren van de resultaten van observationele onderzoeken. Ned Tijdschr Geneesk 2000;144:1393-7.
- 8 The periodic health examination. Canadian Task Force on the Periodic Health Examination. Can Med Assoc J 1979;121:1193-254.
- 9 Everdingen JJE van. Van consensus naar CBO-richtlijn. Ned Tijdschr Geneesk 1999;143:2086-9.
- 10 Assendelft WJJ, Scholten RJPM, Hoving J, Offringa M, Bouter LM. De praktijk van systematische reviews. VIII. Zoeken en beoordelen van systematische reviews. Ned Tijdschr Geneesk 2001;145:1625-31.
- 11 Assendelft WJJ, Scholten RJPM, Eijk JTM van, Bouter LM. De praktijk van systematische reviews. III. Methodologische beoordeling van onderzoeken. Ned Tijdschr Geneesk 1999;143:714-9.
- 12 Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282:1054-60.
- 13 Guyatt GH, Sackett DL, Sinclair JC, Hayward F, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. JAMA 1995;274:1800-4.
- 14 Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. BMJ 1995;311:1356-9.
- 15 Berry G. Statistical significance and confidence intervals. Med J Aust 1986;144:618-9.
- 16 Schuling J, Toenders WGM. Evidence based medicine: betekenis voor de farmacotherapie. Geneesmiddelen Bull 1999;33:71-7.
- 17 Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature. XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. JAMA 1999;282:771-8.
- 18 Scholten RJPM, Kostense PJ, Assendelft WJJ, Bouter LM. De praktijk van systematische reviews. V. Heterogeniteit tussen onderzoeken en subgroepenanalyses. Ned Tijdschr Geneesk 1999;143:843-7.
- 19 Koes BW, Scholten RJ, Mens JM, Bouter LM. Efficacy of epidural steroid injections for low-back pain and sciatica: a systematic review of randomized clinical trials. Pain 1995;63:279-88.
- 20 Watts RW, Silagy CA. A meta-analysis on the efficacy of epidural corticosteroids in the treatment of sciatica. Anaesth Intensive Care 1995;23:564-9.
- 21 Hopayian K, Mugford M. Conflicting conclusions from two systematic reviews of epidural steroid injections for sciatica: which evidence should general practitioners heed? Br J Gen Pract 1999;49:57-61.
- 22 Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. Can Med Assoc J 1997;156:1411-6.
- 23 Ioannidis JPA, Cappelleri JC, Lau J. Issues in comparisons between meta-analyses and large trials. JAMA 1998;279:1089-93.
- 24 Egger M, Smith GD. Misleading meta-analysis. BMJ 1995;310:752-4.
- 25 Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629-34.
- 26 Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996;312:1215-8.
- 27 Kunz F, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomized clinical trials. BMJ 1998;317:1185-90.
- 28 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000;342:1887-9.
- 29 Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000;342:1878-86.
- 30 Zielhuis GA. Dwalingen in de methodologie. XX. Randomisatie als middel tegen verstoring door indicatie. Ned Tijdschr Geneesk 2000;144:1528-31.
- 31 Assendelft WJJ, Zaat JOM, Bouter LM. Systematische reviews als basis voor de NHG-standaard Epicondylitis [abstract]. Ned Tijdschr Geneesk 1997;141:911.
- 32 Cook DJ, Greengold NL, Ellrodt AG, Weingarten SR. The relation between systematic reviews and practice guidelines. Ann Intern Med 1997;127:210-6.
- 33 Assendelft WJJ, Everdingen JJE van. Richtlijnen. In: Offringa M, Assendelft WJJ, Scholten RJPM. Inleiding in evidence-based medicine; klinisch handelen gebaseerd op bewijsmateriaal. Houten: Bohn Stafleu Van Loghum; 2000.
- 34 Feder G, Eccles M, Grol R, Griffiths C, Grimshaw J. Clinical guidelines. Using clinical guidelines. BMJ 1999;318:728-30.

Aanvaard op 16 maart 2002

## Bladvulling

### *De bespreking van een kookboek*

Hoewel voor leken geschreven verdient het werkje ook door die praktizerende geneeskundigen gelezen te worden, wien het aan tijd ontbreekt de omvangrijke handboeken over voedingstherapie van VON LEYDEN, SÉE en anderen te bestuderen, vooral omdat het in een klein bestek een overzicht geeft

van de verteerbaarheid en de voedingswaarde der meest voorkomende spijsen en een aantal kookrecepten bevat, die, hoewel niet voor een hollandsch verhemelte bedacht, toch met geringe moeite daarvoor pasklaar gemaakt kunnen worden.

(Boekaankondigingen. Ned Tijdschr Geneesk 1902;46II:609.)